Computer Vision

Computer Science Tripos Part II

Dr Christopher Town

6. Texture, colour, stereo, and motion descriptors. Disambiguation.

UNIVERSITY OF CAMBRIDGE

Disambiguation

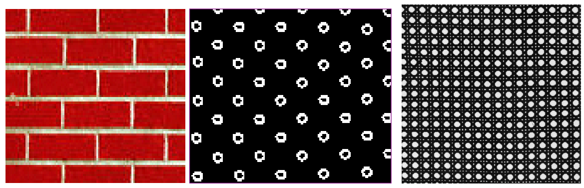
1. The nature, geometry, and wavelength composition of the illuminant(s).
2. Properties of the objects imaged, such as: spectral reflectances; surface shape; surface albedo; surface texture; geometry, motion, and rotation angle.
3. Properties of the camera (or viewer), such as (i) geometry and viewing angle; (ii) spectral sensitivity; (iii) prior knowledge, assumptions, and expectations.

Texture

What defines a texture?

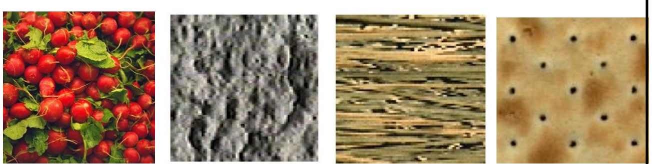
Trevor Darrell

Includes: more regular patterns



Trevor Darrell

Includes: more random patterns



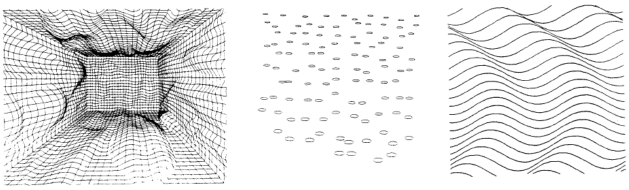
Trevor Darrell

Scale: objects vs. texture

Often the same thing in the world can occur as texture or an object, depending on the scale we are considering.

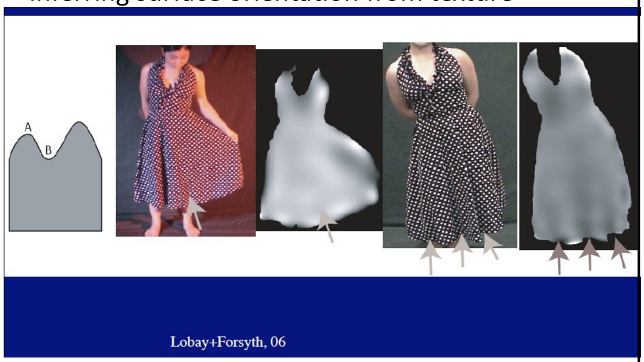
Inferring surface orientation from texture

-> the assumption of **uniformity** constrains the problem



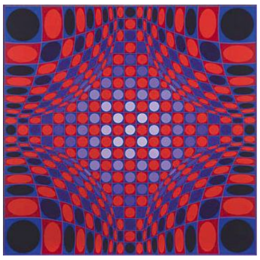
Dr Chris Town

Inferring surface orientation from texture



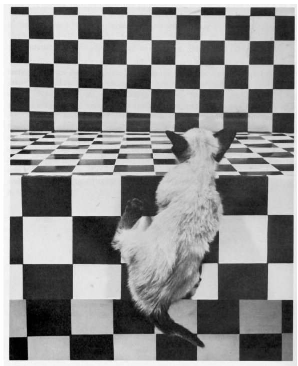
Dr Chris Town

Inferring surface orientation from texture



Dr Chris Town

Inferring surface orientation from texture



The Visual Cliff, by William Vandivert, 1960

Dr Chris Town

Texture

Texture is defined by the existence of certain **statistical correlations** across the image.
Examples:
- quasi-periodic undulations (waves, ripples, folds in clothing)
- spots, speckles
- stripes, dashes

Many natural textures can appear to be almost **fractal**, i.e. **self-similar** across different scales.

The unifying notion in all of these examples is **quasi-periodicity**, or **repetitiveness**, of some features.

Dr Chris Town

So, What Scale to Choose?

- It depends on what we're looking for…

- Too fine a scale… can't see the forest for the trees.
- Too coarse a scale… can't tell the maple from the cherry.

Slide credit: K. Grauman

Dr Chris Town

Texture

- Textures are made up of repeated sub-elements
- Representation:
  - find the sub-elements, and represent their statistics
- But what are the sub-elements, and how do we find and characterise them?

Texture

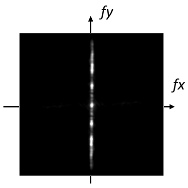
Fourier methods: capture quasi-periodicity at different scales and orientation, but have **non-localised** (global) response

Gabor wavelets: spatially localised, so we can analyse texture in terms of **local spectral analysis**

A simple texture descriptor

Magnitude of the Fourier Transform

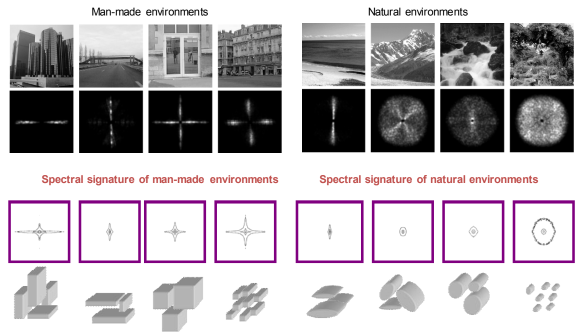
$$A(f_x, f_y) = \left| \sum_{x,y} i(x,y) e^{-2\pi j (f_x x + f_y y)} \right|$$



Magnitude of the Fourier Transform encodes unlocalised information about dominant orientations and scales in the image.

A.Torralba

Statistics of Scene Categories

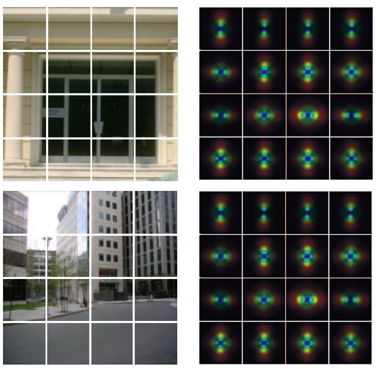


Oliva et al (99), Oliva & Torralba (01)

Look at Mumford's work for models

Gist descriptor

Oliva and Torralba, 2001

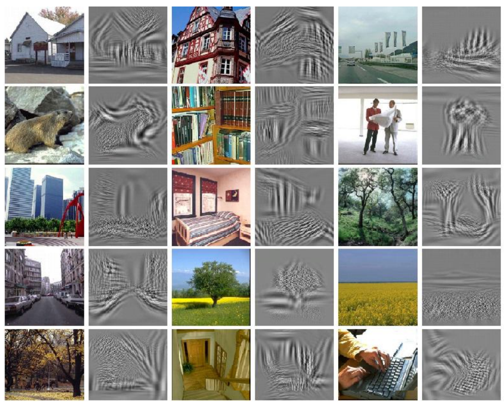


- Apply oriented Gabor filters over different scales
- Average filter energy in each bin

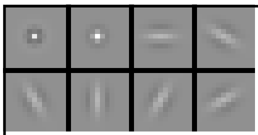
8 orientations
4 scales
x 16 bins
512 dimensions

M. Gorkani, R. Picard, ICPR 1994; Walker, Malik. Vision Research 2004; Vogel et al. 2004; Fei-Fei and Perona, CVPR 2005; S. Lazebnik, et al, CVPR 2006; …

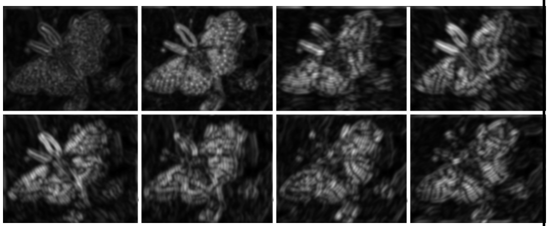
Example visual gists



Global features (I) ~ global features (I')

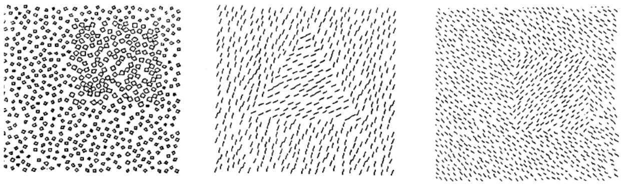


Texture characterisation using filters

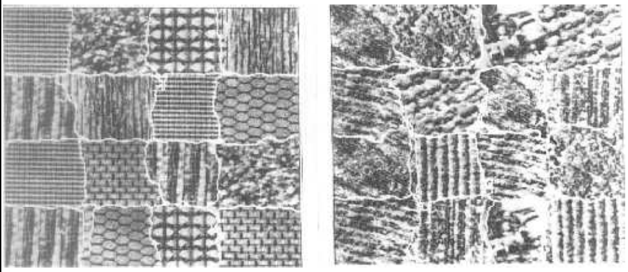


Region segmentation

-> perceptual grouping

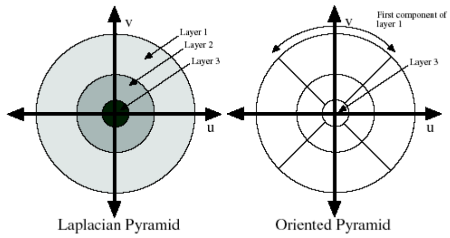


Region segmentation using Gabor Wavelets



Oriented (or "steerable") pyramids

- Laplacian pyramid is orientation independent
- Apply an oriented filter to determine orientations at each layer
  - this represents image information at a particular scale and orientation



Computer Vision - A Modern Approach; D.A. Forsyth

But we need to get rid of the corner regions before starting the recursive circular filtering

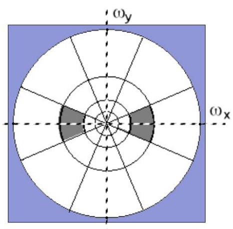
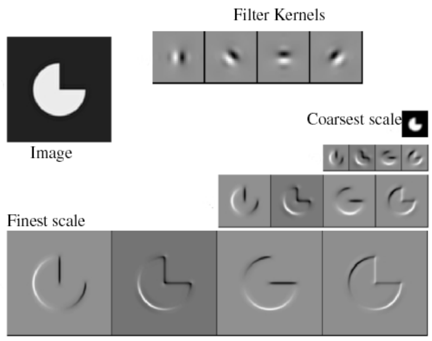


Figure 1. Idealized illustration of the spectral decomposition performed by a steerable pyramid with $k = 4$. Frequency axes range from $-\pi$ to π. The basis functions are related by translations, dilations and *rotations* (except for the initial highpass subband and the final lowpass subband). For example, the shaded region corresponds to the spectral support of a single (vertically-oriented) subband.

http://www.cns.nyu.edu/ftp/eero/simoncelli95b.pdf Simoncelli and Freeman, ICIP 1995



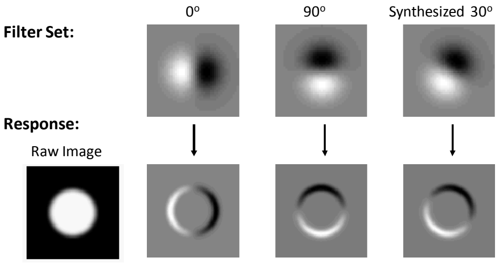
Reprinted from "Shiftable MultiScale Transforms," by Simoncelli et al., IEEE Transactions on Information Theory, 1992, copyright 1992, IEEE

Computer Vision - A Modern Approach; D.A. Forsyth

Steerable filters

"Steerability"-- the ability to synthesise a filter of any orientation from a linear combination of filters at fixed orientations.

The basis functions of the steerable pyramid are **directional derivative** operators, that come in different sizes and orientations.



Taken from:
W. Freeman, T. Adelson, "The Design and Use of Sterrable Filters", IEEE Trans. Patt, Anal. and Machine Intell., vol 13, #9, pp 891-900, Sept 1991

Texture representation

- Form an oriented pyramid (or equivalent set of responses to filters at different scales and orientations).
- Square the output (modulus)
- Take statistics of responses
  - Mean of each filter output (e.g. are there lots of spots?)
  - Standard deviation of each filter output (e.g. are the spots of similar size?)
  - Mean of one scale conditioned on other scale having a particular range of values (e.g. are the spots in straight rows?)

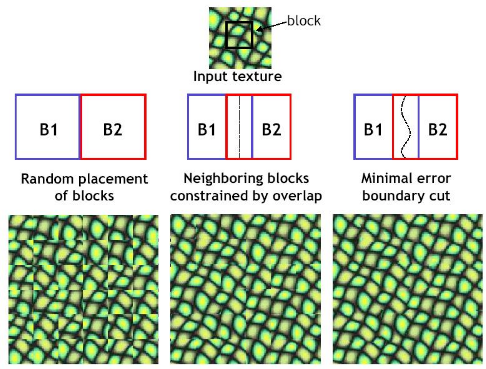
Computer Vision - A Modern Approach; D.A. Forsyth

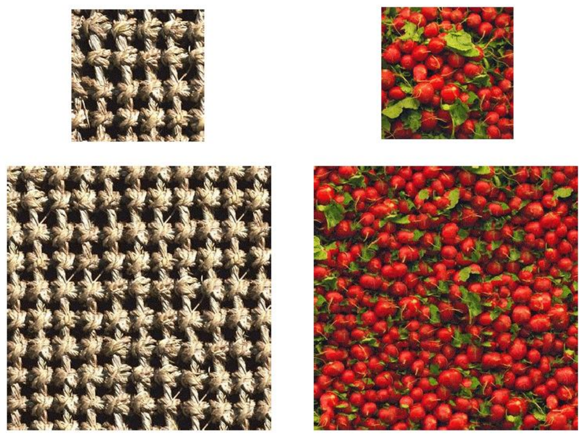
Texture synthesis

- Model texture as generated from random process.
- Discriminate by seeing whether statistics of two processes seem the same.
- Synthesize by generating image with same statistics.

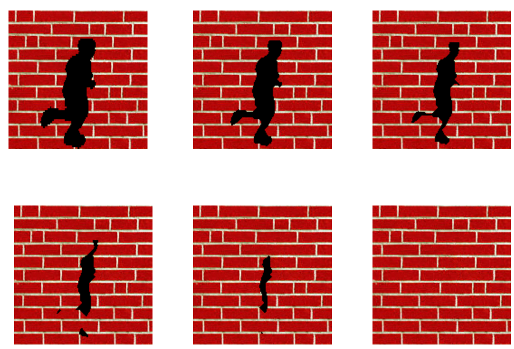
Computer Vision - A Modern Approach; D.A. Forsyth

Texture synthesis



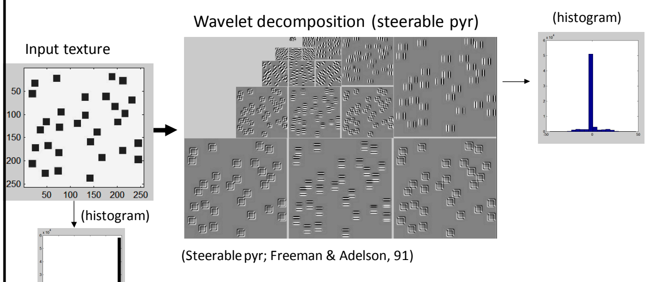


Statistical in-fill



Slide from Alyosha Efros, ICCV 1999

Texture analysis



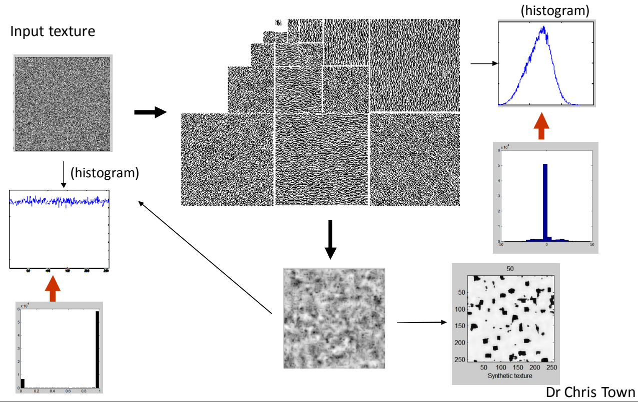
(Steerable pyr; Freeman & Adelson, 91)

The texture is represented as a collection of marginal histograms.

Dr Chris Town

Texture synthesis

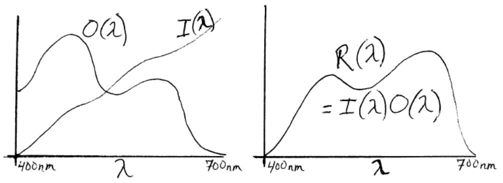
Heeger and Bergen, 1995



Dr Chris Town

Inferring Object Colour

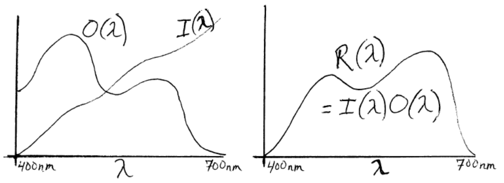
- Let $I(\lambda)$ represent the wavelength composition of the illuminant (i.e. the amount of energy it contains as a function of wavelength λ, across the visible spectrum from about 400 nanometers to 700 nm).
- Let $O(\lambda)$ represent the inherent spectral reflectance of the object at a particular point: the fraction of incident light that is scattered back from its surface there, as a function of the incident light's wavelength λ.
- Let $R(\lambda)$ represent the actual wavelength mixture received by the camera at the corresponding point in the image of the scene.



Dr Chris Town

Inferring Object Colour

Clearly, $R(\lambda) = I(\lambda)O(\lambda)$. The problem is that we wish to infer the "object colour" (its spectral reflectance as a function of wavelength, $O(\lambda)$), but we only know $R(\lambda)$, the actual wavelength mixture received by our sensor. So unless we can measure $I(\lambda)$ directly, how could this problem of inferring $O(\lambda)$ from $R(\lambda)$ possibly be solved?



Dr Chris Town

Inferring Object Colour

Measuring $I(\lambda)$

Search for highly specular (shiny, metallic, glassy) regions in an
Then we could infer ***O*** by dividing ***R*** by ***I***.

Problems:

- We may not find any specular surfaces in the image
- Most materials are not purely specular(e.g. metals which have a brassy colour)
- Not robust, too dependent on highly localised measurements

Dr Chris Town

Colour Constancy

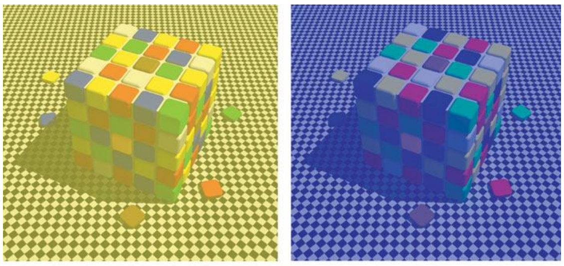
These images show a bowl of fruit photographed in three lighting conditions:
- artificial light (left)
- hazy daylight (middle)
- clear blue sky (right)

Notice the marked variation in colour balance caused by the spectral properties of the illuminant. We are not normally aware of this variation because colour constancy mechanisms discount illumination effects

http://www.psypress.co.uk/mather/resources/topic.asp?topic=ch12-tp-04

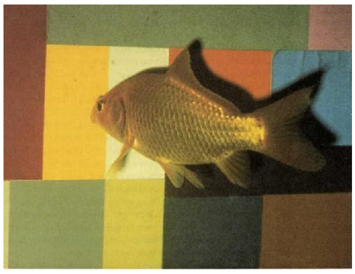
Dr Chris Town

Colour Cube Illusion



D. Purves, R. Beau Lotto, S. Nundy "Why We See What We do," American Scientist

Colour Constancy in Goldfish



In David Ingle's experiment, a goldfish has been trained to swim to a patch of a given color for a reward—a piece of liver. It swims to the green patch regardless of the exact setting of the three projectors' intensities. The behavior is strikingly similar to the perceptual result in humans.

http://neuro.med.harvard.edu/site/dh/b45.htm

Colour Constancy

Possible explanations:

Local colour contrast— cone excitation level of one surface relative to another remains constant when both surfaces experience the same change in illumination. -> Relative cone excitation levels are invariant ratios

Colour adaptation—reduces the contribution from the source illumination by lowering activity in the most highly active cone classes.

Global contrast—global spectral changes generally represent changes in the illuminant; localised differences usually correspond to reflectance differences.

Range of reflected spectrum—gives some indication of the breadth of the illuminating spectrum.

Colour constancy is not perfect (83% accuracy), and the most powerful cue to constancy is local **colour contrast**.

http://www.psypress.co.uk/mather/resources/topic.asp?topic=ch12-tp-04

Retinex

E.H. Land, J.J. McCANN - Journal of the Optical society of America, 1971

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 61, NUMBER 1 JANUARY 1971

Lightness and Retinex Theory

EDWIN H. LAND* AND JOHN J. MCCANN
Polaroid Corporation, Cambridge, Massachusetts 02139
(Received 8 September 1970)

Retinex

The key idea is that *the colours of objects or areas in a scene are determined by their surrounding spatial context.* A complex sequence of ratios computed across all the boundaries of objects (or areas) enables the illuminant to be algebraically discounted in the sense shown in the previous Figure, so that object spectral reflectances $O(\lambda)$ which is what we perceive as their colour, can be inferred from the available retinal measurements $R(\lambda)$ without explicitly knowing $I(\lambda)$.

Retinex

Reflectance tends to be constant across space except for abrupt changes at the transitions between objects.

Thus a reflectance change shows itself as step edge in an image, while illuminance changes gradually over space.

By this argument one can separate reflectance change from illuminance change by measuring the response to spatial derivatives.

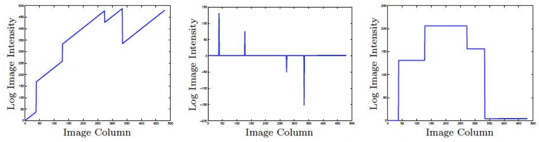
Retinex

Again, we are trying to solve an ill-posed problem

From M. Tappen,

Dr Chris Town

Retinex

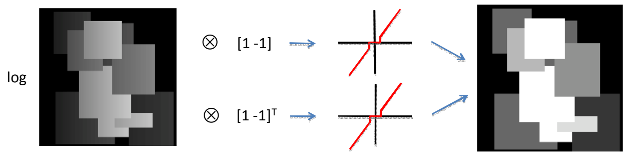


(a) One column from the observed image.

(b) The derivative of the plot from (a).

(c) The estimate of the log shading

From M. Tappen, PhD

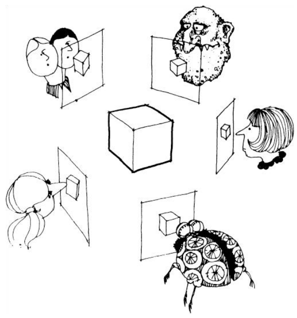


From M. Tappen,

Dr Chris Town

What Is Stereo Vision?

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape

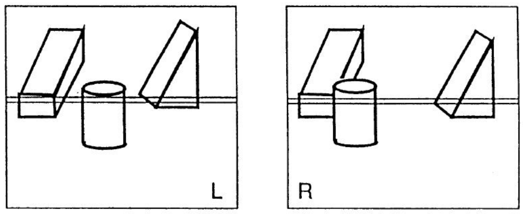


Slide credit: Svetlana Lazebnik, Steve Seitz

Dr Chris Town

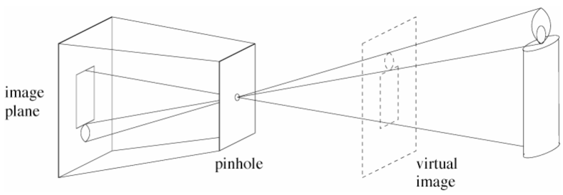
Stereo

Important information about depth can be obtained from the use of two (or more) cameras, in the same way that humans achieve stereoscopic depth vision by virtue of having two eyes. Objects in front or behind of the point in space at which the two optical axes intersect (as determined by the angle between them, which is controlled by camera movements or eye movements), will project into different relative parts of the two images. This is called *stereoscopic disparity.*



Dr Chris Town

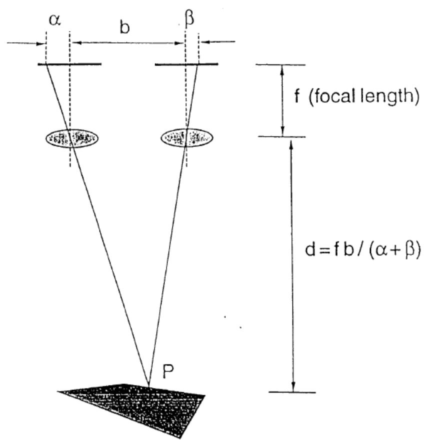
Pinhole Camera



Source: Forsyth & Ponce

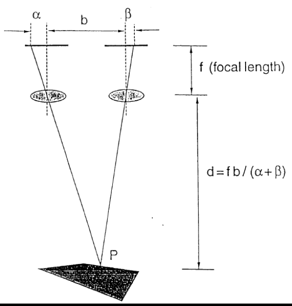
47

Dr Chris Town



Dr Chris Town

This "error signal" becomes greater in proportion to the distance of the object in front or behind the point of fixation, and so it can be calibrated to obtain a depth cue. It also becomes greater with increased spacing between the two eyes or cameras, since that is the "base of triangulation." (That is why WWI armies introduced V-shaped binocular "trench periscopes" to increase stereoscopic visual acuity, for breaking camouflage by increasing the *effective* spacing between the viewer's two eyes to almost a meter.)



Dr Chris Town

Camera calibration

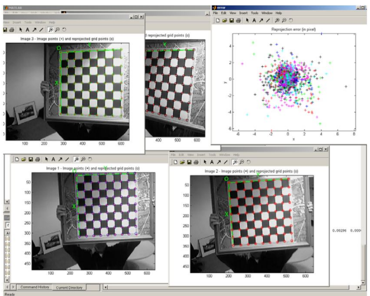
Camera parameters:
- 6 degrees-of-freedom in space (3 spatial coordinates X,Y,Z and 3 Euler rotation angles)
- Focal length.

Relative orientation:
- Base of separation
- Alignment in space
- Difference in focal length
- Rotation around each camera's optical axis

Dr Chris Town

Camera calibration

There are many other possible ***intrinsic*** camera calibration parameters, such as
- **skew coefficients** accounting for non-orthogonality
- **distortion coefficients** representing radial and tangential distortions of the lens



Dr Chris Town

The Correspondence Problem

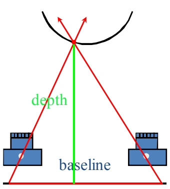
Features (pixels, edge responses, SIFT features etc.) in the two images need to be matched

If each image has N features, then there are **N^2** possible pairings

However, the number of potential pairings is
N x **(N-1)** x **(N-2) … = N!**

Dr Chris Town

Stereo vision



Triangulate on two images of the same point to recover depth.
- Feature matching across views
- Calibrated cameras

Matching correlation windows across scan lines

Slide credit: David Lowe

Dr Chris Town

Simplest Case: Rectified Images

- Image planes of cameras are parallel.
- Focal points are at same height.
- Focal lengths same.
- Then, epipolar lines fall along the horizontal scan lines of the images

- We will assume images have been *rectified* so that epipolar lines correspond to scan lines
  - Simplifies algorithms
  - Improves efficiency

Slide credit: David Lowe

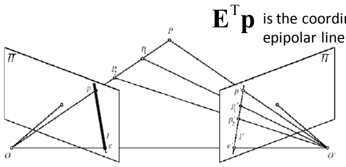
Dr Chris Town

Essential Matrix and Epipolar Lines

$\boxed{\mathbf{p}'^{\mathrm{T}}\mathbf{E}\mathbf{p} = 0}$

Epipolar constraint: if we observe point p in one image, then its position p' in second image must satisfy this equation.

$\mathbf{E}^{\mathrm{T}}\mathbf{p}$ is the coordinate vector representing the epipolar line for point p

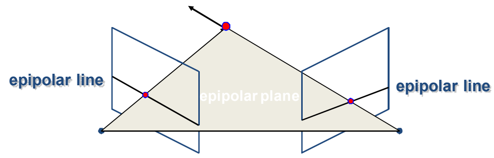


$\mathbf{E}\mathbf{p}'$ is the coordinate vector representing the epipolar line for point p'

Slide credit: K. Grauman

Dr Chris Town

The epipolar constraint



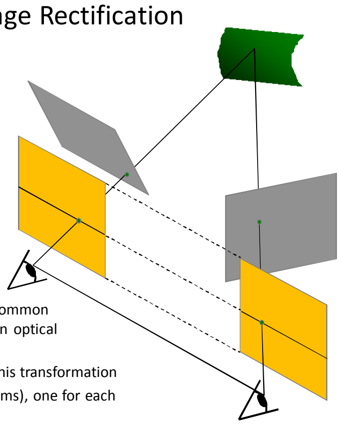
- Epipolar Constraint
  - Matching points lie along corresponding epipolar lines
  - Reduces correspondence problem to 1D search along *conjugate epipolar lines*
  - Greatly reduces cost and ambiguity of matching

Slide credit: David Lowe

Dr Chris Town

Recap: Stereo Image Rectification

- In practice, it is convenient if image scanlines are the epipolar lines.



- Algorithm
  - Reproject image planes onto a common plane parallel to the line between optical centers
  - Pixel motion is horizontal after this transformation
  - Two homographies (3x3 transforms), one for each input image reprojection

C. Loop & Z. Zhang, Computing Rectifying Homographies for Stereo Vision, CVPR'99

Dr Chris Town

Motion information

- For stereo vision, we need to solve the Correspondence Problem for two images simultaneous in time but acquired with a spatial displacement.
- For motion vision, we need to solve the Correspondence Problem for two images coincident in space but acquired with a temporal displacement.
- The object's spatial "disparity" can be measured in the two image frames once their backgrounds have been aligned. This can be calibrated to reveal motion information when compared with the time interval, or depth information when compared with the binocular spatial interval.

Dr Chris Town

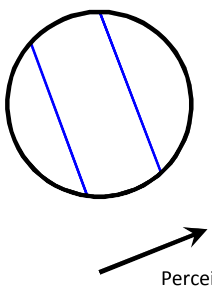
Motion information

Among the challenging requirements of motion detection and inference are:

1. Need to infer 3D object trajectories from 2D image motion information.
2. Need to make *local* measurements of velocity, which may differ in different image regions in complex scenes with many moving objects. Thus, a velocity vector field needs to be assigned over an image.
3. Need to disambiguate object motion from contour motion, so that we can measure the velocity of an object regardless of its *form*.
4. Need to measure velocities regardless of the size of the viewing aperture in space and in time (the spatial and temporal integration windows). This is known as the *aperture problem*.
5. It may be necessary to assign more than one velocity vector to any given local image region (as occurs in "motion transparency")
6. We may need to detect a *coherent* overall motion pattern across many small objects or regions separated from each other in space.

Dr Chris Town

The Aperture Problem

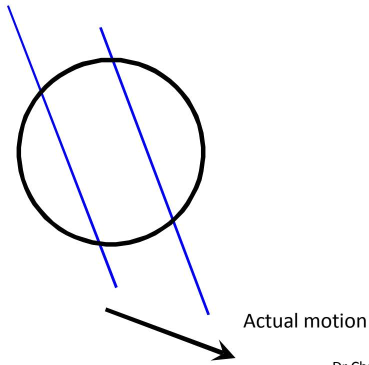


Perceived motion

Slide credit: Svetlana Lazebnik

Dr Chris Town

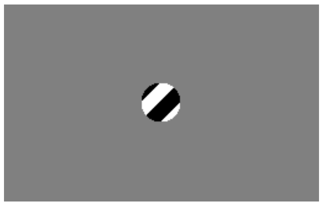
The Aperture Problem



Slide credit: Svetlana Lazebnik

Dr Chris Town

The Barber Pole Illusion



http://en.wikipedia.org/wiki/Barberpole_illusion

Slide credit: Svetlana Lazebnik

Dr Chris Town

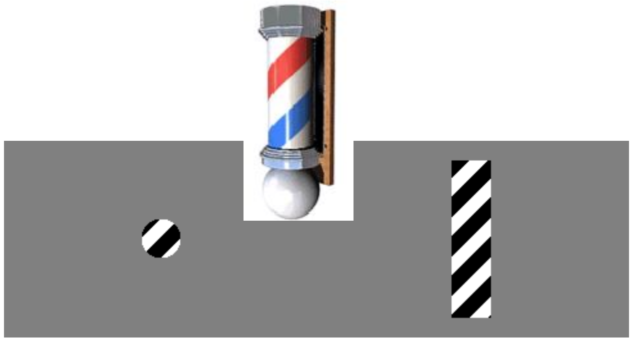
The Barber Pole Illusion

http://en.wikipedia.org/wiki/Barberpole_illusion

Slide credit: Svetlana Lazebnik

Dr Chris Town

The Barber Pole Illusion



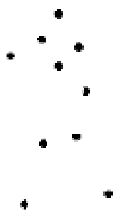
Dr Chris Town

- Automated motion analysis generally limited to opaque and solid objects
- Challenges: flocks of birds, clouds, vapours, waves, fire, the wind in the willows…

Dr Chris Town

Motion and Perceptual Organisation

- Even "impoverished" motion data can evoke a strong percept

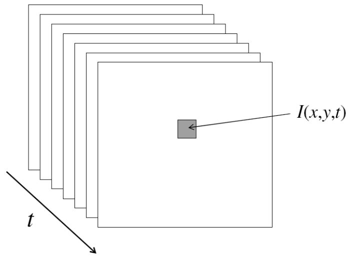


Slide credit: Svetlana Lazebnik

Dr Chris Town

Video

- A video is a sequence of frames captured over time
- Now our image data is a function of space (x, y) and time (t)



Slide credit: Svetlana Lazebnik

Dr Chris Town

Motion Estimation Techniques

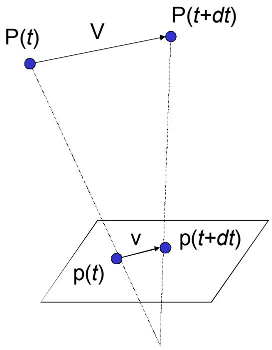
- Direct methods
  - Directly recover image motion at each pixel from spatio-temporal image brightness variations
  - Dense motion fields, but sensitive to appearance variations
  - Suitable for video and when image motion is small
- Feature-based methods
  - Extract visual features (corners, textured areas) and track them over multiple frames
  - Sparse motion fields, but more robust tracking
  - Suitable when image motion is large (10s of pixels)

Slide credit: Steve Seitz

Dr Chris Town

Motion Field and Parallax

- $P(t)$ is a moving 3D point
- Velocity of scene point: $V = dP/dt$
- $p(t) = (x(t),y(t))$ is the projection of P in the image.
- Apparent velocity v in the image: given by components $v_x = dx/dt$ and $v_y = dy/dt$
- These components are known as the *motion field* of the image.

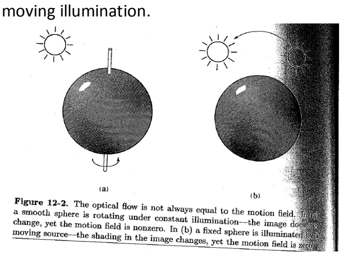


Slide credit: Svetlana Lazebnik

Dr Chris Town

Optical Flow

- Definition: optical flow is the *apparent* motion of brightness patterns in the image.
- Ideally, optical flow would be the same as the motion field.
- Have to be careful: apparent motion can be caused by lighting changes without any actual motion.
  - Think of a uniform rotating sphere under fixed lighting vs. a stationary sphere under moving illumination.



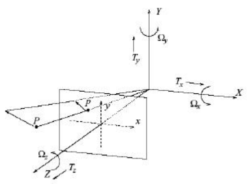
Slide credit: Svetlana Lazebnik

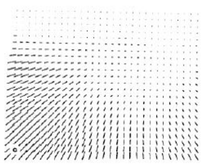
Dr Chris Town

Optical Flow Field

Image irradiance at time t and location $\mathbf{x}=(x, y)$

$$I(x,y,t)$$





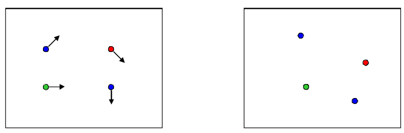
$u(x,y)$ Horizontal component

$v(x,y)$ Vertical component

M. Black

Dr Chris Town

Estimating Optical Flow



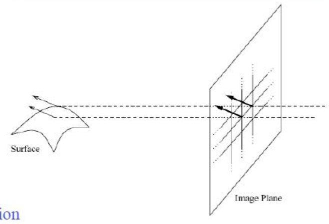
$I(x,y,t-1)$ $I(x,y,t)$

- Given two subsequent frames, estimate the apparent motion field *u*(*x*,*y*) and *v*(*x*,*y*) between them.
- Key assumptions
  - Brightness constancy: projection of the same point looks the same in every frame.
  - Small motion (temporal coherence): points do not move very rapidly.
  - Spatial coherence: points move like their neighbors.

Slide credit: Svetlana Lazebnik

Dr Chris Town

Spatial Coherence



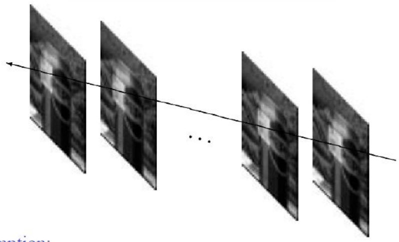
Assumption

* Neighboring points in the scene typically belong to the same surface and hence typically have similar motions.
* Since they also project to nearby points in the image, we expect spatial coherence in image flow.

M. Black

Dr Chris Town

Temporal Persistence



Assumption:

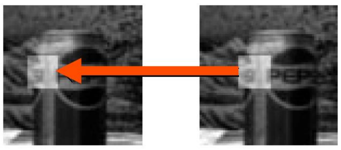
The image motion of a surface patch changes gradually over time.

M. Black

Dr Chris Town

Optical flow constraint (also known as *Brightness constancy constraint*)

Brightness Constancy



$$I(x+u, y+v, t+1) = I(x, y, t)$$

(assumption)

Dr Chris Town

The Brightness Constancy Constraint

$$I(x+u, y+v, t+1) = I(x, y, t)$$

(assumption)

Taylor Series Expansion

$dx = u,\ dy = v,\ dt = 1$

Assume u, v, dt small

Assume brightness varies *smoothly* with x, y, t

$$I(x,y,t) + dx\frac{\partial}{\partial x}I(x,y,t) + dy\frac{\partial}{\partial y}I(x,y,t) + dt\frac{\partial}{\partial t}I(x,y,t) + \varepsilon$$

M Black

Dr Chris Town

The Brightness Constancy Constraint

$$I(x+u, y+v, t+1) = I(x, y, t)$$

$$I(x+u, y+v, t+1) - I(x, y, t) = 0$$

$$\cancel{I(x,y,t)} + dx\frac{\partial}{\partial x}I(x,y,t) + dy\frac{\partial}{\partial y}I(x,y,t) + dt\frac{\partial}{\partial t}I(x,y,t) - \cancel{I(x,y,t)} = 0$$

Divide through by dt

$$u\frac{\partial}{\partial x}I(x,y,t) + v\frac{\partial}{\partial y}I(x,y,t) + \frac{\partial}{\partial t}I(x,y,t) = 0$$

$$I_x u + I_y v + I_t = 0$$

M Black

Dr Chris Town

Notation

$$I_x u + I_y v + I_t = 0$$

$$\nabla I^T \mathbf{u} = -I_t$$

$$\nabla I = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix}$$

M Black

Dr Chris Town

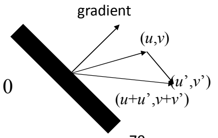
The Brightness Constancy Constraint

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

- How many equations and unknowns per pixel?
  - One equation, two unknowns
- Intuitively, what does this constraint mean?

$$\nabla I \cdot (u, v) + I_t = 0$$

- The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown



If (u,v) satisfies the equation, so does $(u+u', v+v')$ if $\nabla I \cdot (u', v') = 0$

Slide credit: Svetlana Lazebnik

79 Dr Chris Town

Intensity Gradient Models

Assume that the local time-derivative in image intensities at a point, across many image frames, is related to the local spatial gradient in image intensities because of object velocity $\vec{v}$:

$$-\frac{\partial I(x,y,t)}{\partial t} = \vec{v} \cdot \vec{\nabla} I(x,y,t)$$

Then the ratio of the local image time-derivative to the spatial gradient is an estimate of the local image velocity (in the direction of the gradient).

Dr Chris Town

Dynamic Zero-Crossing Models

Measure image velocity by first finding the edges and contours of objects (using the zero-crossings of a blurred Laplacian operator!), and then take the time-derivative of the Laplacian-Gaussian-convolved image:

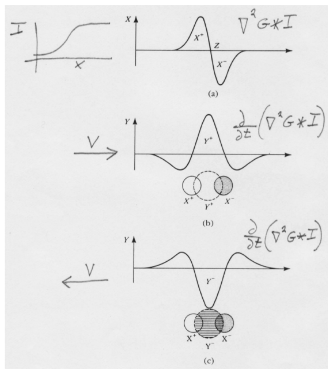
$$-\frac{\partial}{\partial t}\left[\nabla^2 G_\sigma(x,y) * I(x,y,t)\right]$$

in the vicinity of a Laplacian zero-crossing. The amplitude of the result is an estimate of speed, and the sign of this quantity determines the direction of motion relative to the normal to the contour.

Also known as the "Hildreth model", after Ellen Hildreth

Dr Chris Town

Dynamic Zero-Crossing Models

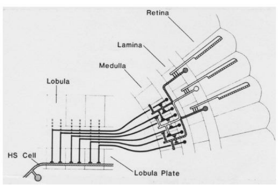


Dr Chris Town

Spatio-Temporal Correlation Models

Image motion is detected by observing a *correlation* of the local image signal $I(x,y,t)$ across an interval of space and and after an interval of time τ. Finding the pair of these intervals which maximises the correlation between $I(x,y,t)$ and $I(x - v_x\tau, y - v_y\tau, t - \tau)$ determines the two components of image velocity v_x and v_y which we desire to know.

$$\text{argmax} \int_x \int_y \int_t I(x,y,t) \cdot I(x - v_x\tau,\ y - v_y\tau,\ t - \tau)\ dx\ dy\ dt$$



Detailed studies of fly neural mechanisms (above) for motion detection and visual tracking led to elaborated correlation-based motion models.

Dr Chris Town

Spatio-Temporal Spectral Models

It is possible to detect and measure image motion purely by *Fourier* means. This approach exploits the fact that motion creates a covariance in the spatial and temporal *spectra* of the time-varying image $I(x,y,t)$, whose three-dimensional (spatio-temporal) Fourier transform is defined:

$$F(\omega_x, \omega_y, \omega_t) = \int_X \int_Y \int_T I(x,y,t) e^{-i(\omega_x x + \omega_y y + \omega_t t)} dx dy dt$$

In other words, rigid image motion has a 3D spectral consequence: the local 3D spatio-temporal spectrum, rather than filling up 3-space $(\omega_x, \omega_y, \omega_t)$, collapses onto a 2D inclined plane which includes the origin. Motion detection then occurs just by filtering the image sequence in space and in time, and observing that tuned spatio-temporal filters whose centre frequencies are **co-planar** in this 3-space are activated together. This is a consequence of the **Spectral Co-Planarity Theorem**, which states that translational image motion of velocity $\vec{\mathbf{v}}$ has a 3D spatio-temporal Fourier spectrum that is non-zero only on an inclined plane through the origin of frequency-space. Spherical coordinates of the unit normal to this spectral plane correspond to the speed and direction of motion.

Dr Chris Town

Theorem: Translational image motion of velocity $\vec{\mathbf{v}}$ has a 3D spatio-temporal Fourier spectrum that is non-zero only on an inclined plane through the origin of frequency-space. Spherical coordinates of the unit normal to this spectral plane correspond to the speed and direction of motion.

Let $I(x, y, t)$ be a continuous image in space and time.

Let $F(\omega_x, \omega_y, \omega_t)$ be its 3D spatio-temporal Fourier transform:

$$F(\omega_x, \omega_y, \omega_t) = \int_X \int_Y \int_T I(x, y, t) e^{-i(\omega_x x + \omega_y y + \omega_t t)} dx dy dt.$$

Let $\vec{\mathbf{v}} = (v_x, v_y)$ be the local image velocity.

Uniform motion $\vec{\mathbf{v}}$ implies that for all time shifts t_o,

$$I(x, y, t) = I(x - v_x t_o, y - v_y t_o, t - t_o).$$

Taking the 3D spatio-temporal Fourier transform of both sides, and applying the shift theorem, gives

$$F(\omega_x, \omega_y, \omega_t) = e^{-i(\omega_x v_x t_o + \omega_y v_y t_o + \omega_t t_o)} F(\omega_x, \omega_y, \omega_t).$$

Dr Chris Town

Taking the 3D spatio-temporal Fourier transform of both sides, and applying the shift theorem, gives

$$F(\omega_x, \omega_y, \omega_t) = e^{-i(\omega_x v_x t_o + \omega_y v_y t_o + \omega_t t_o)} F(\omega_x, \omega_y, \omega_t).$$

The above equation can only be true if $F(\omega_x, \omega_y, \omega_t) = 0$ everywhere the exponential term doesn't equal 1.

This means $F(\omega_x, \omega_y, \omega_t)$ is non-zero only on the 3D spectral plane

$$\boxed{\omega_x v_x + \omega_y v_y + \omega_t = 0} \qquad \textbf{Q.E.D.}$$

The spherical coordinates $(\theta, \phi, 1)$

$$\phi = \tan^{-1}\left(\omega_t / \sqrt{\omega_x^2 + \omega_y^2}\right)$$

$$\theta = \tan^{-1}(\omega_y / \omega_x)$$

of the inclined spectral plane's unit normal are determined by $\vec{\mathbf{v}}$ and correspond to the *speed* (ϕ) and *direction* (θ) of motion:

$$\phi = \sqrt{v_x^2 + v_y^2}$$

$$\theta = \tan^{-1}(v_y / v_x)$$

Dr Chris Town